

ZUMA-Forschung zur Methodenentwicklung: Bericht über das Projekt "Befragungsexperimente"

Hoppler, Hans-Jürgen; Trometer, Reiner; Schwarz, Norbert

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Hoppler, H.-J., Trometer, R., & Schwarz, N. (1983). ZUMA-Forschung zur Methodenentwicklung: Bericht über das Projekt "Befragungsexperimente". *ZUMA Nachrichten*, 7(12), 4-30. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210516>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ZUMA-FORSCHUNG ZUR METHODENENTWICKLUNG: BERICHT ÜBER DAS PROJEKT "BEFRAGUNGSEXPERIMENTE"

1. Einleitung

Standardwerke der Umfrageforschung, die Ende der 40er Jahre bis Mitte der 50er Jahre in den USA erschienen sind (CANTRIL, 1944; HYMAN, 1954; KAHN & CANNELL, 1957) und derzeit immer noch Gültigkeit besitzen, enthalten bereits kritische Reflexionen über die methodische Qualität angewandter Erhebungs- und Auswertungstechniken. Trotz einer erheblichen Zunahme an standardisierten Befragungen und einer technischen Verfeinerung des Instrumentariums gibt es bis heute keine Theorie der Befragung, die den Interaktions- und Urteilsprozessen im Interview Rechnung trägt und die Einführung genereller Hypothesen über Störfaktoren erlaubt, um die methodische Qualität empirischer Untersuchungen fundiert zu gewährleisten.

IRLE (1981) spricht in diesem Zusammenhang von einem "scientific lag" in der Methodik der Umfrageforschung, der sich zunehmend vergrößere. In den letzten Jahren hat nun - nach zeitweiliger Beschränkung auf Untersuchungen zum Einfluß der Interviewer auf Befragungen - vor allem in den USA eine Renaissance kritischer Überprüfungen der generellen Methodik der Umfrageforschung eingesetzt (SCHUMAN & PRESSER, 1981; DIJKSTRA & VAN DER ZOUWEN, 1982; SUDMAN & BRADBURN, 1982). Dabei wird z. B. die Entwicklung von Fragebögen weniger als "Kunst" betrachtet, deren Grundlage nur in langer "Erfahrung" erworben werden kann (NOELLE, 1963; MAYNTZ et al., 1971; SCHEUCH, 1973), sondern eher als systematische Aufarbeitung empirischer Befunde, um damit die verschiedenen Störquellen bei der Befragung besser eingrenzen zu können. Diese Versuche einer Kodifizierung der Einflußfaktoren im Interview erfordert allerdings eine interdisziplinäre Vorgehensweise: "Umfrageforschung kann nur insoweit ein dominanter Weg sozialwissenschaftlicher Forschung sein, indem sie an gegenwärtiger Forschung in den Nachbarwissenschaften lernt" (IRLE, 1981).

Diese Überlegungen haben bei ZUMA dazu geführt, die Forschungen zur Methodenentwicklung in diesem Bereich zu intensivieren, um die Qualität der Einzelberatung externer Forschungsprojekte ständig weiter zu verbessern.

Bei diesen Forschungen handelt es sich sowohl um die Auswertung spezifischer Fragebogenteile (z. B. "Kontaktprotokolle" beim ALLBUS 1980) und die Sekundäranalyse einschlägiger Daten in ausgewählten Repräsentativuntersuchungen (z. B. WOHLFAHRTSSURVEY 1980, ZUMABUS 6) als auch um speziell experimentell ausgerichtete Studien.

Unter dem Gesichtspunkt der Verbesserung der laufenden Dienstleistungen im Umfragebereich erscheint es angezeigt, der besonderen Untersuchungssituation im Feld Rechnung zu tragen: Schwerpunktmäßig werden quasi-experimentelle Versuchspläne eingesetzt (COOK & CAMPBELL, 1979), die, im Gegensatz zu im Labor durchgeführten Experimentalplänen, zwar keine lückenlose, so doch eine annähernd vollständige Kontrolle der unabhängigen Variablen ermöglichen.

Hierüber soll in diesem Beitrag berichtet werden. Wir geben dabei zunächst einen Überblick über die verschiedenen Fragestellungen, denen wir in diesen Studien nachgehen wollen, um dann an zwei ausgewählten Beispielen Vorgehensweise und erste Ergebnisse exemplarisch zu diskutieren.

2. Beschreibung der Fragestellungen

Die einzelnen Fragestellungen, die im Rahmen des Projekts "Befragungsexperimente" genauer untersucht werden sollen, betreffen die folgenden Problemkreise:

A. Ausschöpfung.

In den letzten Jahren wurde die Umfrageforschung zunehmend mit dem Problem der Ausschöpfung der vorgesehenen Stichprobe konfrontiert. Deshalb wird die Wirkung unterschiedlicher Anreize auf die Teilnahmebereitschaft an Panelstudien ebenso untersucht wie auch der Ausschöpfungsgrad in Abhängigkeit von der soziodemographischen Struktur einzelner Stadtquartiere, Anschreibeversionen oder Interviewermerkmalen. Erste Ergebnisse deuten darauf hin, daß Interviewermerkmalen eine entscheidende Bedeutung bei unterschiedlichen Ausschöpfungsraten zukommt. In einem weiteren Schritt ist geplant, diese Problematik durch Sekundäranalysen anhand ausgewählter Repräsentativbefragungen näher einzugrenzen.

B. Fragebogenumfang und Fragenabfolge.

Ein zweiter Bereich gilt der Untersuchung von Einflüssen des Fragebogenumfangs und der Fragenreihenfolge auf die "Antwortqualität" und die Motivation bzw. Aufmerksamkeit des Befragten. In einer Teiluntersuchung wurde eine Variation der Länge und des Inhalts des Fragebogens im bisher als "kritisch" angesehenen Bereich von etwa einer Stunde Befragungszeit vorgenommen. Erste Ergebnisse weisen bisher lediglich auf eine veränderte Beurteilung des Instruments durch die Interviewer hin: Bei einer längeren Befragungszeit wird die Interviewsituation von den Interviewern - jedoch nicht von den Befragten - negativ beurteilt.

C. Fragebogenkonstruktion und Wortwahl.

In einem dritten Bereich sind bereits im Rahmen der laufenden Forschungsarbeiten methodische Replikationen amerikanischer Arbeiten durchgeführt worden, und zwar einmal zur Vorfilterung "Meinungsloser" und zum anderen zum "verbieten/erlauben"-Phänomen (SCHUMAN & PRESSER, 1981). Durch diese Untersuchungen soll zunächst überprüft werden, ob die jeweiligen Effekte interkulturell replizierbar sind, um dann in weiteren Arbeiten die jeweiligen Ursachen dieser Effekte zu bestimmen.

D. Unangenehme bzw. heikle Fragen.

In Umfragen beschäftigt uns immer wieder das Problem der "besten" Konstruktion "unangenehmer" bzw. "heikler" Fragen (KOOLWIJK, 1968, 1969). In diesem Kontext wurde unter Einbezug sozialpsychologischer Erkenntnisse ("foot-in-the-door"- und "door-in-the-face"-Phänomen; DE JONG, 1979; CANN et al., 1975; SNYDER & CUNNINGHAM, 1975) eine Frage zur Einkommensnennung konstruiert, die zu einer erheblichen Verminderung von Verweigerungen führte. Das Instrument wurde bisher in zwei Umfragen mit einer Spezialstichprobe (Erwerbstätige) und in einer repräsentativen Bevölkerungsumfrage (ZUMABUS 6) eingesetzt. Ein ausführlicher Bericht hierzu ist in Arbeit.

E. Frageformulierung und Antwortverhalten.

Ein nicht unerheblicher Teil methodischer Untersuchungen zu "Qualität" von Umfragedaten befaßt sich mit sogenannten "verzerrten Fragen" (FRIEDRICHS, 1973: 198), d. h. dem Einfluß unterschiedlicher Frageformulierungen bei gleichen Frageinhalten auf das Antwortverhalten der Befragten.

Mehrere Einzeluntersuchungen sollen zeigen, inwieweit variierte Frageformulierungen veränderte Antwortverteilungen bedingen.

F. Kontexteffekte.

Durch unterschiedliche Aktivierung gesellschaftlicher Wertvorstellungen werden Effekte auf nachfolgende politische Einstellungs- und Verhaltensmessungen geprüft (HIPPLER & STRACK, 1981). Das Interesse gilt den in der praktischen Umfrageforschung mehrfach thematisierten "Kontexteffekten" (BRADBURN & MASON, 1964). Unsere Erklärungsgrundlage stützt sich hierbei auf Ergebnisse der sogenannten "Salience-Forschung", in denen u. a. nachgewiesen werden konnte, daß Urteile von Individuen von der kognitiven Verfügbarkeit ("availability") urteilsrelevanter Informationen beeinflusst werden (TVERSKY & KAHNEMAN, 1973).

G. Definitionseffekte bei Einstellungsskalen.

Anhand neuerer Arbeiten der kognitiven Psychologie zum Einfluß aktivierter Information (WYER, 1974; WYER & CARLSTON, 1979) wird in einer weiteren Teilstudie untersucht, welchen Einfluß das vorherige Nachdenken der Befragten über die Bedeutung der Skalenetikette auf die nachfolgenden Einstellungsberichte hat.

H. Einfluß kategorialer Antwortvorgaben.

Schließlich soll der Frage nachgegangen werden, welchen Einfluß unterschiedliche kategoriale Antwortvorgaben auf Verhaltensberichte und nachfolgende Einstellungsmessungen haben. Über erste Ergebnisse hierzu wird im Schlußteil dieser Arbeit berichtet.

Bislang sind folgende Studien durchgeführt worden:

Zwei mündliche Befragungen (zweistufiges Panel) einer zufällig ausgewählten Stichprobe Mannheimer wahlberechtigter Bürger (1. Welle: realisiertes N = 113 / 2. Welle: realisiertes N = 85); im Herbst fand bei einer neuen Zufallsstichprobe wahlberechtigter Mannheimer Bürger eine telefonische Kurzbefragung statt (N = 109). Für April 1983 ist eine weitere mündliche Befragung anhand von Quoten mit einem realisierten N von 150 Befragten vorgesehen.

In den folgenden Abschnitten 3 und 4 stellen wir erste Ergebnisse zu zwei Teilfragestellungen des Gesamtprojekts vor. In Abschnitt 3 geht es um den Aspekt der Fragenkonstruktion und der Formulierung. Dabei wird durch die methodische Replikation zur Vorfilterung "Meinungsloser" und zum "verbieten/erlauben"-Phänomen an bisherige Forschungstraditionen angeknüpft, um in der Diskussion der Ergebnisse alternative Erklärungshypothesen und Vorschläge zu weiteren Arbeiten zu bieten. In Abschnitt 4 geht es um den Einfluß kategorialer Antwortvorgaben auf Verhaltensberichte und nachfolgende Einstellungsmessungen. Dabei wird auf der Grundlage theoretischer Überlegungen eine praktische Fragestellung der Umfrageforschung experimentell überprüft, um damit die Verbindung zwischen psychologischer Theorie und praxisnahen Problemen herzustellen.

3. Fragenkonstruktion und Formulierung: Methodische Replikationen zur Vorfilterung "Meinungsloser" und zum "verbieten/erlauben"-Phänomen

Trotz wesentlich verbesserten Instrumentariums zur statistischen Datenanalyse, trotz häufig geforderter Untersuchungen zur Klärung der "psychologischen" Bedingungen des Befragungsprozesses (etwa: NOELLE-NEUMANN, 1970; SCHUMAN & PRESSER, 1977), stehen wir immer noch am Anfang einer umfassenden "Theorie der Befragung" (KAHN & CANNELL, 1957). Effekte der Formulierung und der Konstruktionsmerkmale von Fragen werden in gängigen deutschsprachigen Lehrbüchern der empirischen Sozialforschung (vgl. z. B. HOLM, 1975; KOOLWIJK & WIEKEN-MAYSER, 1974; SCHEUCH, 1973) mit einigen Ausnahmen (ESSER, 1975; ERBSLÖH et al., 1973) meist nur am Rande behandelt. In einer neuen amerikanischen Veröffentlichung (SCHUMAN & PRESSER, 1981) wurde jetzt erstmals eine Vielzahl einzelner Studien zusammengefaßt und damit der Versuch unternommen, Gemeinsamkeiten und Regelmäßigkeiten der unterschiedlichen Effekte bei der Formulierung und Konstruktion von Fragen herauszuarbeiten. Die generelle Interpretationslinie des Buches der beiden Autoren ist jedoch auf eine Klärung des Einflusses der gefundenen Effekte auf die jeweils untersuchten soziologischen Tatbestände ausgerichtet. Populations-spezifika wie Bildung, Alter oder Geschlecht etc. werden herangezogen, um die gefundenen Phänomene aufzuklären. In den nachfolgenden Experimenten gilt unser Interesse jedoch ausschließlich der Frage der Übertragbarkeit von zwei bei SCHUMAN und PRESSER ausführlich dokumentierten Effekten. Bei der Diskussion beider methodischer Replikationen zur Vorfilterung "Mei-

nungsloser" und zum "verbieten/erlauben"-Phänomen wird darauf verzichtet, einen Zusammenhang zwischen gefundenen Effekten und soziologischen Erklärungsvariablen herzustellen, weil unseres Erachtens in Folgeuntersuchungen zunächst geklärt werden muß, welche kognitiven Prozesse diesen Effekten zugrunde liegen.

3.1 Experiment I: Zur Vorfilterung "Meinungsloser"

3.1.1 Spezifikation der Fragestellung

Jeder, der Umfragedaten erhebt oder damit arbeitet, wird spätestens bei der Auswertung mit der Schwierigkeit konfrontiert, sogenannten "nicht inhaltlichen" Antworten einen bestimmten Stellenwert innerhalb der Untersuchung zuzuordnen. In den Standardprozeduren der meisten Computer-Software-Programme werden solche Antworten als nicht real vorhandene Codes (Missing Data) gezählt, die dann im weiteren Verlauf der Analyse meist unberücksichtigt bleiben.

Das Problem, wie "weiß nicht"-Angaben behandelt werden sollen, stellt sich jedoch bereits bei der Konstruktion des Fragebogens. Die meist geübte Praxis besteht darin, "weiß nicht" als Antwortmöglichkeit bei bestimmten Einzelfragen zwar zuzulassen, aber die Interviewer im voraus so zu instruieren, daß sie die Nennung dieser Kategorie nur dann aufzeichnen, wenn sie spontan den vorgegebenen inhaltlichen Antwortalternativen vorgezogen wird. In manchen Fällen allerdings, besonders bei Einstellungsfragen, ist diese Antwortmöglichkeit unerwünscht und daher im Fragebogen auch nicht vorgesehen, so daß es letztlich den individuellen Präferenzen des Interviewers überlassen bleibt, ob er die "weiß nicht"-Äußerung eines Befragten zuläßt und abweichend von den Kategorienvorgaben notiert, oder ihn drängt, sich für eine der inhaltlichen Vorgaben zu entscheiden.

Gegen diese Behandlung der "weiß nicht"-Antworten als quantité négligeable wenden sich mittlerweile eine ganze Reihe amerikanischer Autoren, die in ihren Untersuchungen nachweisen konnten, daß solche "nicht inhaltlichen" Befragtenangaben keineswegs, wie häufig unterstellt, zufällig verteilt sind und deshalb auch nicht einfach ausgeklammert oder negiert werden können. FRANCES und BUSH (1975) zeigen beispielsweise, daß bestimmte Bevölkerungs-

gruppen (Niedriggebildete, Personen mit niedrigem Einkommen, Frauen), die von Informationen, Einfluß und Entscheidungsprozessen ausgeschlossen sind, eine erhöhte Anzahl von "weiß nicht"-Antworten aufweisen. Auch FERBER (1966) macht auf eine systematische Beziehung zwischen "nicht inhaltlichen" Befragtenangaben einerseits und Geschlecht bzw. Alter, Bildungsgrad sowie Haushaltsstrukturvariablen andererseits aufmerksam.

In Anlehnung an SCHUMAN und PRESSER (1981) sollen in dem hier geschilderten Experiment die Effekte zweier unterschiedlicher "weiß nicht"-Handhabungen (Vorfilterform und Standardform) auf die Anzahl von "weiß nicht"-Äußerungen sowie die Verteilung der inhaltlichen Antworten untersucht werden. Unter einer "Vorfilterform" verstehen wir, daß der Befragte, bevor er eine Stellungnahme zu einem bestimmten Thema abgeben soll, danach gefragt wird, ob er dazu überhaupt eine Meinung habe. "Standardform" bedeutet in diesem Zusammenhang, daß die "weiß nicht"-Kategorie als Antwortmöglichkeit bei der Kategorienvorgabe für den Interviewer zwar erscheint, jedoch von ihm nur dann anzukreuzen ist, wenn der Befragte offensichtlich eine Frage weder befürworten noch ablehnen kann.

3.1.2 Durchführung

85 Befragten wurde in der zweiten Welle der vorgestellten Studie zwei Fragen zur Problematik der persönlichen Freiheit in der Bundesrepublik gestellt. Der einen Hälfte der Befragten wurden die beiden einzelnen Fragen vorgelesen, und sie wurden jeweils direkt im Anschluß daran nach Zustimmung bzw. Ablehnung gefragt, wobei die "weiß nicht"-Kategorie lediglich für den Interviewer in dem Fall vorgesehen war, daß der Befragte den Inhalt der Frage weder befürworten noch ablehnen konnte (Standardform). Die andere Hälfte der Befragten bekam jeweils zuerst die Frage ohne Antwortvorgaben vorgelesen, um im Anschluß daran gefragt zu werden, ob sie zu diesem Satz eine Meinung hätten; bejahten sie diese Frage, wurden sie danach um ihre Zustimmung bzw. Ablehnung gebeten (Vorfilterform).

Untenstehend ein Beispiel zu beiden Frageformen:

<p>Standardform</p> <p>A Ich habe jetzt zwei Fragen zur Einschränkung der persönlichen Freiheit in der Bundesrepublik:</p> <p>Stimmen Sie dem folgenden Satz zu oder lehnen Sie ihn ab?</p> <p>Durch das Verhalten der Polizei bei Demonstrationen wird der Freiheitsspielraum des einzelnen Bürgers immer mehr eingeengt.</p>		
stimme zu	1	
lehne ab	2	
<hr/> <p>weiß nicht 8</p>		

<p>Vorfilterform</p> <p>A Ich habe jetzt zwei Fragen zur Einschränkung der persönlichen Freiheit in der Bundesrepublik. Nicht jeder hat eine Meinung zu diesen Fragen. Uns interessiert jetzt zuerst, ob Sie eine Meinung zu dem jeweiligen Satz haben und erst anschließend, ob Sie ihm zustimmen oder ob Sie ihn ablehnen.</p> <p>Durch das Verhalten der Polizei bei Demonstrationen wird der Freiheitsspielraum des einzelnen Bürgers immer mehr eingeengt.</p> <p>Haben Sie zu diesem Satz eine Meinung?</p>		
ja	1	B
nein	2	C
<p>B Stimmen Sie diesem Satz zu oder lehnen Sie ihn ab?</p> <p>INT.: Evtl. nochmals Satz wiederholen.</p>		
stimme zu	1	
lehne ab	2	

3.1.3 Ergebnisse

Die Verteilungen der Antworten auf beide Fragen in der Standardform und der Vorfilterform sind in Tabelle 1 dargestellt.

Tab. 1: Verteilung der Antworten (Standardform, Vorfilterform)

Frage 1: "Durch das Verhalten der Polizei bei Demonstrationen wird der Freiheitsspielraum des einzelnen Bürgers immer mehr eingeengt."

<u>Standardform</u>		<u>Vorfilterform</u>	
stimme zu	18.8%	stimme zu	27.0%
lehne ab	77.1%	lehne ab	59.5%
weiß nicht	4.1%	weiß nicht	13.5%
	100.0%		100.0%
	(48)		(37)

Frage 2: "Man muß heute mehr als früher aufpassen, was man öffentlich sagt oder was man unterschreibt."

<u>Standardform</u>		<u>Vorfilterform</u>	
stimme zu	43.8%	stimme zu	43.2%
lehne ab	56.2%	lehne ab	35.1%
weiß nicht	--	weiß nicht	21.6%
	100.0%		100.0%
	(48)		(37)

Bei dem ersten Item erhöht sich der Anteil der "weiß nicht"-Antworten von der Standard- zur Vorfilterform um 9.4%. Setzt man diesen Unterschied allerdings in Bezug zu den substantiellen Antworten, so ist er nicht so bedeutend ($\chi^2(1)=2.4$, n.s.) wie bei dem zweiten Item: Hier unterscheiden sich die "weiß nicht"-Antworten bei beiden Formen um 21.6%. Dieser Unterschied ist deutlich signifikant, $\chi^2(1)=11.5$, $p < .001$ und zeigt, daß durch die Vorfilterung ein Anstieg der "Meinungslosen" bis zu einem Viertel der Befragten zu beobachten ist. Es ist bei diesem Ergebnis zu vermuten, daß die jeweilige "Zentralität" des Frageinhalts für die Befragten, also die unmittelbare "Bedeutsamkeit" bzw. die "kognitive Repräsentanz" (zum Überblick: LIPPERT & WAKENHUT, 1978) ebenso für die Angabe "keine Meinung" zu

haben eine Rolle spielt wie die "Reserviertheit" gegenüber den Fragen allgemein (LEVERKUS-BRÜNING, 1966) bzw. die "Unangenehmheit" der Fragen (VON KOOLWIJK, 1969). Durch die Standardfrageform werden die vom Befragten unter diesen Gesichtspunkten zu erwartenden Reaktionen der Meinungslosigkeit allerdings größtenteils unterdrückt, während die Vorfilterform hierzu offensichtlich eher Ausdrucksmöglichkeiten zuläßt.

Die praktische Frage, die sich hieraus ergibt, ist, ob der hohe Anteil Meinungsloser bei der vorgefilterten Version des zweiten Items die inhaltliche Verteilung der Frage selbst verändert. Hierfür werden die "weiß nicht"-Antworten als "missing data" behandelt und die prozentuale Verteilung neu berechnet:

Tab. 2: Antwortverteilung bei Nicht-Berücksichtigung der "weiß nicht"-Antworten

Frage 2: "Man muß heute mehr als früher aufpassen, was man öffentlich sagt oder was man unterschreibt."

<u>Standardform</u>		<u>Vorfilterform</u>	
stimme zu	43.8%	stimme zu	55.2%
lehne ab	56.2%	lehne ab	44.8%
	100.0%		100.0%
	(48)		(29)

Durch die Herausnahme der "weiß nicht"-Antworten verschieben sich die Verteilungen bei den beiden Formen der zweiten Frage nur minimal, $\chi^2(1)=1.0$, n.s., da lediglich ein konstanter Faktor eliminiert wurde. Inhaltlich läßt sich zwischen den Verteilungen der Standardform und der Vorfilterform jedoch ein bedeutsamer Unterschied feststellen: Die am häufigsten gewählte Kategorie ist bei der Standardform die Ablehnung des Items, bei der Vorfilterform jedoch die Zustimmung. In Anbetracht der kleinen Fallzahl ist dieser Unterschied jedoch nicht statistisch signifikant. Bei SCHUMAN und PRESSER (1981) waren in 14 von 19 Experimenten keine inhaltlichen Unterschiede zwischen beiden Frageformen festzustellen, und vier der fünf signifikanten Differenzen erwiesen sich als relativ gering.

3.1.4 Diskussion

Die Ergebnisse der methodischen Replikation zur unterschiedlichen Handhabung von "weiß nicht"-Antworten bestätigen weitgehend die interkulturelle Übertragbarkeit der von SCHUMAN und PRESSER (1981) gefundenen Effekte. Durch die Methode der Vorfilterung kann es im Unterschied zur Standardmethode zu einem 20-25% höheren Anteil von "weiß nicht"-Antworten kommen.

Vergleicht man die durch beide Methoden erhaltenen Verteilungen, lassen sich inhaltlich Unterschiede ausmachen, die allerdings in unserem Beispiel nicht statistisch signifikant waren. SCHUMAN und PRESSER (1981) fanden darüber hinaus (anhand der größeren Stichproben möglich) die Beziehungen zwischen den Antworten und demographischen Hintergrundvariablen relativ wenig verändert; jedoch wurden die Beziehungen der Einstellungsfragen untereinander in einigen Fällen durch die Art des Filters beeinflusst, allerdings sowohl in Richtung stärkerer als auch schwächerer Assoziationen.

Der bei der Vorfilterung auftretende Anteil von "weiß nicht"-Antworten sollte uns dennoch dazu anregen, über die inhaltliche Qualität der Antworten, die wir ohne Vorselektion "Meinungsloser" in Umfragen erzielen, intensiver nachzudenken. CONVERSE (1970) konnte in diesem Zusammenhang auf beeindruckende Weise zeigen, daß sich das Problem der "Nichttattitüden" bei Einstellungsfragen ohne die vorherige Selektion der Befragten durch Filterfragen noch wesentlich schwerwiegender darstellt. Der Anteil der "weiß nicht"-Antworten hängt außerdem von der Art des Filters selbst ab (BISHOP et al., 1978). Die "Nichtwissenden" bilden somit keinen "harten Kern", sondern variieren in ihrer Zusammensetzung je nach Wortwahl und Konstruktion des Filters. Die Erklärungskonzepte, die bei den meisten Arbeiten zu dieser Problematik bisher herangezogen worden sind (Wortwahl- und Konstruktionsmerkmale von Filtern/Befragtenmerkmalen), sind nicht ausreichend. Es ist vielmehr notwendig, dem Informationsaufnahme- und Verarbeitungsprozeß des Befragten im Augenblick der Fragestellung erhöhte Aufmerksamkeit zu widmen, um hierdurch zu einer theoretisch fundierten Modellannahme zum "Nichtwissen" bei Einstellungsmessungen zu gelangen.

3.2 Experiment II: Zum "verbieten/erlauben"-Phänomen

3.2.1 Spezifikation der Fragestellung

Immer wieder sind wir bei der Konstruktion von Einstellungsfragen mit dem Problem der Frageformulierung befaßt. Warnungen, wie sie etwa DAVIS (1976) ausspricht, daß schon kleine Veränderungen der Frageformulierung starke Einflüsse auf die Antwortverteilungen zeigen, werden dabei allzuoft mit dem Argument zurückgewiesen, daß diese Effekte größtenteils individuell bei einzelnen Fragen auftreten, nicht ohne weiteres generalisierbar sind und auf eine Vielzahl von Fragen angewendet werden können.

Dieses Argument ist jedoch in dieser Form nicht stichhaltig, weil einerseits generell zu wenig (Vor-)Untersuchungen zur Wortwahl bei Einstellungsfragen durchgeführt werden und man sich andererseits nur wenig um die experimentelle Variation einzelner Grundmerkmale bei Frageformulierungen auf der Grundlage theoriegeleiteter Prozeßannahmen kümmert (etwa: "Ja-sage-Tendenz"; "Einfluß von Mittelkategorien" etc.). Bereits 1941 fand RUGG in einem Experiment, daß die Unterstützung freier Meinungsäußerung um etwa 1/5 höher ausfiel, wenn die Frage lautete, "Do you think the United States should forbid public speeches against democracy?", als wenn "forbid" durch "allow" ersetzt wurde. Drei Replikationen von SCHUMAN und PRESSER (1981) in den 70iger Jahren bestätigten diesen Befund: Amerikanische Befragte sind eher bereit, die freie Meinungsäußerung nicht zu erlauben als zu verbieten, obgleich beide Vorgaben logisch äquivalent sind. In weiteren Untersuchungen konnte dann gezeigt werden, daß dieser Effekt auch bei anderen, konkreteren Frageinhalten (Pornographie, Zigarettenwerbung) auftritt.

Im folgenden Experiment wollen wir deshalb diesen über Jahrzehnte in amerikanischen Untersuchungen stabilen Frageformulierungseffekt methodisch replizieren, d. h. unabhängig vom inhaltlichen Kontext in der deutschen Adaption nachzuweisen versuchen, um anschließend die sich hieraus ergebenden Forschungskonsequenzen zu diskutieren.

3.2.2 Durchführung

Den Befragten der vorangegangenen Untersuchung (85 zufällig ausgewählte

wahlberechtigte Mannheimer) wurde eine Frage zu sogenannten "Peep-Shows" (zum "bezahlten Sich-zur-Schau-Stellen nackter Frauen in aufreizenden Posen") gestellt. Die eine Hälfte der Probanden wurde hierbei gefragt, ob der Gesetzgeber diese "Peep-Shows" generell erlauben solle, bei der anderen Hälfte wurde das Wort "erlauben" durch "verbieten" ersetzt. Um Unterschiede in der Argumentationsstruktur bei den beiden Vorgaben zu untersuchen, wurden die Befragten anschließend noch mit einer offenen Frage um eine Begründung ihrer Meinung gebeten.

3.2.3 Ergebnisse

Der von SCHUMAN und PRESSER (1981) gefundene Effekt konnte bestätigt werden (vgl. Tabelle 3).

Tab. 3: Antwortverteilung ("verbieten/erlauben")

<u>inhaltliche Antwort</u>	<u>Frageversion</u>	
	verbieten	erlauben
"Peep-Shows" verbieten (nicht erlauben)	27.7%	50.0%
"Peep-Shows" erlauben (nicht verbieten)	72.3%	50.0%
	100.0% (47)	100.0% (38)

Bei der "verbieten"-Version der Frage sind 22.3% der Befragten mehr für ein "Erlauben" von Peep-Shows als in der "erlauben"-Version, $\chi^2(1)=3.5$, $p < .06$, d. h. umgekehrt "nicht zu erlauben" fällt den Befragten von der Formulierung her leichter als zu "verbieten".

Bei der geforderten Begründung der Entscheidung durch die Befragten produzieren die beiden Vorgaben bei den Befürwortern bzw. Gegnern von "Peep-Shows" keine Unterschiede im Inhalt ihrer Argumentation pro bzw. contra. Durchschnittlich 50% der Befragten, die "Peep-Shows" verboten sehen möchten, argumentieren bei den Vorgaben mit "moralischen Bedenken" und "der Würde der Frau"; 35% mit "Gefahren für Frauen bzw. Jugend" und mit "wachsender Kriminalität" (15% sonstige Argumente).

Durchschnittlich 75% der Befragten, die für ein Erlauben von "Peep-Shows" plädieren, ziehen Argumente wie "persönliche Freiheit", "Demokratie" oder "Recht auf Privatleben" für die Begründung ihrer Entscheidung unabhängig vom vorgegebenen Stimulus "verbieten" oder "erlauben" heran (25% sonstige Argumente).

3.2.4 Diskussion

Die Ergebnisse zeigen, daß in einer methodischen Replikation mit verändertem Frageinhalt der von SCHUMAN und PRESSER (1981) berichtete Effekt im interkulturellen Vergleich nachweisbar ist.

Die nachfolgende Analyse der Inhalte der Entscheidungsbegründung (pro bzw. contra Peep-Shows) zeigte keine interpretierbaren Unterschiede. SCHUMAN und PRESSER (1981) sehen drei Erklärungsmöglichkeiten:

- Zum einen wird angenommen, daß die jeweiligen Begriffe (verbieten/erlauben) eine unterschiedliche grammatikalische Satzstruktur erzeugen;
- dann wird vermutet, daß der starke Effekt bei der Replikation des RUGG-Items - "Meinungsäußerung gegen die Demokratie" - daher rührt, daß der Inhalt für die Befragten nicht zentral bzw. zu abstrakt war;
- schließlich wird angenommen, daß die unterschiedliche Konnotation von verbieten bzw. erlauben dazu führt, daß die Befragten sich eher gegen den "härteren" Stimulus "verbieten" entscheiden.

Die Überprüfung dieser post-hoc-Erklärungen führt bei SCHUMAN und PRESSER (1981) außer einer Interaktion mit Bildungsvariablen zu keiner befriedigenden Lösung: "We are hard put to see any obvious source of the effect" (S. 296). Unseres Erachtens sind jedoch weitaus weniger komplizierte Erklärungen notwendig, um die Ursachen dieses Effekts einzugrenzen. Zu vermuten ist, daß die Gleichsetzung von "verbieten" mit "nicht erlauben" und von "erlauben" mit "nicht verbieten", die durch die einseitige (und in Umfragen normalerweise nicht übliche) Fragestellung künstlich erzeugt wird, so nicht angebracht ist. Wir nehmen vielmehr an, daß von der Begrifflichkeit her ein Antwortkontinuum existiert, das von "verbieten" über "nicht erlauben" und

"nicht verbieten" zu "erlauben" je nach Fragestellung Pro- bzw. Kontra-Positionen unterschiedlicher Stärke miteinschließt. Mit "verbieten" werden vermutlich nur diejenigen Befragten erfaßt, die aktiv dagegen sind, mit "nicht erlauben" die gleiche Gruppe, aber vermutlich zusätzlich diejenigen, die der Problematik indifferent gegenüberstehen; gleiches gilt im umgekehrten Sinn für "erlauben": Hier ist die Gruppe der aktiven Befürworter vertreten, wohingegen "nicht verbieten" wiederum vermutlich indifferente Befragte miteinschließt. Diese Annahme soll mit Hilfe eines externen Rating-Verfahrens der Einzelargumente in einer bisher allerdings nur im Planungsstadium befindlichen Untersuchung geklärt werden.

4. Untersuchungen zum Einfluß kategorialer Antwortvorgaben auf Verhaltensberichte und nachfolgende Einstellungsmessungen

Die Erfassung (quasi-)objektiver Daten erfolgt in der Umfrageforschung häufig durch die Vorgabe einer Anzahl von Kategorien, auf denen der Befragte sich einordnen soll, um so z. B. über seinen Fernsehkonsum, seine Kirchengangshäufigkeit o. a. zu berichten. Unsere Kenntnisse über die kognitiven Prozesse, die einer solchen Selbstbeschreibung durch Kategorienauswahl zugrunde liegen, sind ebenso lückenhaft wie der Wissensstand über den Einfluß, den eine vom Befragten vorgenommene Festlegung auf eine spezifische Antwort auf nachfolgende verwandte Fragen haben mag.

Bei der Konstruktion von Antwortkategorien zur Frequenzmessung versuchen Umfrageforscher in der Regel, den vermuteten Frequenzbereich möglichst vollständig abzudecken. Zusätzlich gehen in die Konstruktion der Kategorienvorgabe Überlegungen zur einfachen Handhabbarkeit der Vorgabe für den Befragten sowie zur Auswertungsökonomie mit ein. Schließlich empfehlen einige Autoren, Kategorien, die in besonderer Weise "sozial erwünscht" sind, ans Ende der Vorgabe zu setzen, da andernfalls die Befragten häufiger die anfangs aufgeführte, sozial erwünschte Kategorie nutzen würden, ohne die weiteren Kategorien überhaupt zu berücksichtigen (SUDMAN & BRADBURN, 1982). Über solche fragebogentechnische Erwägungen hinaus sollte jedoch bedacht werden, daß Skalenvorgaben die Urteilsprozesse der Befragten in vielfältiger Weise zu beeinflussen vermögen, wie dies für Ratingskalen (UPSHAW, 1969; SHERMAN et al., 1978) und Rangordnungsskalen (SCHWARZ, WYER & KRUG- LANSKI, 1982) in verschiedenen Untersuchungen gezeigt wurde. Ziel der vor-

liegenden Arbeit ist es, einige Einflüsse kategorialer Antwortvorgaben auf Verhaltensberichte und nachfolgende Urteile aufzuzeigen, sowie vorläufige Konzipierungen der diesen Einflüssen zugrunde liegenden Prozesse vorzutragen. Insbesondere möchten wir mit den nachfolgend berichteten Experimenten zeigen, daß die Frage in ihrer Gesamtheit, also Fragetext und Antwortvorgabe, als eine Einheit zu betrachten ist, die einerseits den Befragten Informationen über den Fragegegenstand und die Erwartungen des Forschers vermittelt und andererseits je nach Kategorienvorgabe unterschiedliche Urteilsprozesse beim Befragten initiiert.

Betrachtet man die Information, die Fragetext und Kategorienvorgabe vermitteln, so zeigt sich, daß Befragte im allgemeinen anzunehmen scheinen, daß Forscher ihre Skalen so konstruieren, daß die Mehrzahl der Befragten den Mittelbereich der Skala besetzt. Erste Untersuchungen hierzu (berichtet in SCHWARZ & STRACK, 1982) zeigten, daß Studenten z. B. vermuteten, der "durchschnittliche Student" habe mehr Freunde, wenn die Kategorienvorgabe einer ihnen vorgelegten Frage von "4 und weniger" in Zweierschritten bis "10 und mehr", als wenn sie von "0" bis "4 und mehr" reichte. Sofern Personen aus der Reichweite der Kategorienvorgabe Rückschlüsse auf die wahrscheinliche reale Verteilung ziehen, können sie diesen Rückschlüssen Vergleichsinformationen entnehmen, die sowohl ihre eigene Zuordnung zu diesen Kategorien wie spätere verwandte Urteile zu beeinflussen vermag. Diese Überlegungen werden im ersten Experiment aufgegriffen und geprüft.

Über diese Informationsvermittlung hinaus ist zu vermuten, daß die unterschiedliche Breite der Kategorienvorgabe die Detailliertheit der Erinnerung eigenen Verhaltens beim Befragten beeinflußt. Die Vorgabe einer kurzen Kategorienfolge erfordert eine begrenztere und weniger detaillierte Erinnerung des zu berichtenden Verhaltens, als wenn es sich um eine ausführliche Kategoriengruppe handelt. Darüber hinaus ist zu vermuten, daß durch die unterschiedliche Aktivierung und den Bericht von Verhaltensinformationen ebenfalls nachfolgende Urteile beeinflußt werden - eine Möglichkeit, die in Experiment II exploriert wird.

4.1 Experiment I: Einflüsse kategorialer Antwortvorgaben auf Verhaltensbeurichte

4.1.1 Spezifikation der Fragestellung

Im ersten Experiment wurde der durchschnittliche tägliche Fernsehkonsum der Befragten mit einer offenen Frage oder mit einer von zwei Listenvorgaben erfaßt. Eine Liste reichte von "bis 1/2 Stunde" in Halbstundenschritten bis zu "2 1/2 Stunden und mehr", während die zweite Liste bei "bis 2 1/2 Stunden" begann und mit "4 1/2 Stunden und mehr" endete.

In einer Voruntersuchung wurden die Listen je 10 Beurteilern vorgelegt, die die vermutete Antwort des "Durchschnittsbürgers" schätzen sollten. Wie erwartet, wurde dem Durchschnittsbürger bei Vorlage der hohen Kategorienvorgabe ein größerer Konsum zugeschrieben als bei Vorlage der niederen Vorgaben. Sofern die Befragten eine ähnliche Inferenz ziehen, sollte dies ihren Bericht des eigenen Fernsehkonsums beeinflussen. Darüber hinaus sollte die durch die Kategorienvorgabe vermittelte Vergleichsinformation auch Urteile der Befragten über ihre Freizeitzufriedenheit beeinflussen, vermutlich derart, daß Personen, die erschließen, daß sie mehr fernsehen als der Durchschnittsbürger, mit ihrer Freizeitgestaltung unzufriedener sind als Personen, die erschließen, daß sie weniger fernsehen.

4.1.2 Durchführung

In der zweiten Welle der vorgestellten Untersuchung wurde 85 Personen mit einer offenen Frage (diese diente zur Kontrolle des durchschnittlichen Fernsehkonsums der Stichprobe) oder einer der oben beschriebenen Listenvorgaben die Frage gestellt: "Wie viele Stunden sehen Sie an einem normalen Wochentag durchschnittlich fern?" Dieser Frage schloß sich eine weitere nach der Zufriedenheit mit der Freizeitgestaltung an, die mit einer 10-Punkte-Skala (10 = sehr zufrieden / 1 = sehr unzufrieden) erhoben wurde.

4.1.3 Ergebnisse

Tabelle 4 gibt den berichteten Fernsehkonsum wieder.

ZUMA

Tab. 4: Berichteter Fernsehkonsum

S K A L A 1			S K A L A 2			
bis	1/2 Stunde	34%				
1/2	bis 1 Stunde	14%				
1	Stunde bis 1 1/2 Stunden	10%				
1 1/2	Stunden bis 2 Stunden	21%				
2	Stunden bis 2 1/2 Stunden	14%	bis 2 1/2	Stunden	63%	
2 1/2	Stunden und mehr	7%	2 1/2	Stunden bis 3	Stunden	22%
			3	Stunden bis 3 1/2	Stunden	4%
			3 1/2	Stunden bis 4	Stunden	7%
			4	Stunden bis 4 1/2	Stunden	4%
			4 1/2	Stunden und mehr	--	
		100%			100%	
		(29)			(27)	

OFFENE FRAGE:

minimal = 1/2 Stunde / maximal = 4 Stunden / MITTELWERT = 2.0 Stunden (N=27)

Bei der offenen Frage lag der mittlere Fernsehkonsum, wie erwartet, im Bereich von 2 Stunden. Dieser mittlere Konsum wird demnach bei Skala 1 mit dem Endbereich, bei Skala 2 mit dem Anfangsbereich abgedeckt. Dennoch sind die Verteilungen der beiden Skalen (Anfangskategorie am stärksten, Endkategorie am geringsten besetzt) relativ gleichartig und ergeben inhaltlich unterschiedliche Berichte des Fernsehkonsums, wie die Reduktion beider Skalenversionen auf ihren Schnittbereich zeigt (vgl. Tabelle 5).

Tab. 5: Berichteter Fernsehkonsum (Schnittbereich Skala 1/Skala 2)

Fernsehkonsum	Skala 1	Skala 2
bis 2 1/2 Stunden	93%	63%
2 1/2 Stunden und mehr	7%	37%
	100%	100%
	(29)	(27)

Wie erwartet, berichteten mehr Personen bei der niederen als bei der hohen Kategorienvorgabe einen Konsum bis 2 1/2 Stunden, $\chi^2(1)=9.7$, $p < .01$. Das Ergebnis legt nahe, daß für Personen mit hohem Fernsehkonsum der Bericht

dieses Konsums auf der Skala mit hoher Vorgabe erleichtert wurde, während bei geringen Vorgaben ein hoher Konsum als unüblich erschien und nicht durch Besetzung der undifferenzierten Endkategorie berichtet wurde.

Konsistent mit diesen Vergleichsprozeßüberlegungen wurde auch die Zufriedenheit der Befragten mit der Freizeitgestaltung durch die Listenvorgabe beeinflusst. Vergleicht man die Befragten, die einen Konsum von bis zu 2 1/2 Stunden berichten (die Daten der anderen Befragten sind nicht vergleichbar), zeigt sich wie erwartet, daß Personen, denen die Liste mit hohen Vorgaben vorlag und die daraus vermutlich schlossen, daß andere mehr fernsehen als sie selbst, höhere Zufriedenheit mit ihrer Freizeit berichteten als Personen, denen die Liste mit niedrigen Vorgaben vorlag. Der Mittelwertvergleich liefert einen t-Wert von 1.8 (34), $p < .09$ (vgl. Tabelle 6).

Tab. 6: Vergleich der Freizeitzufriedenheit

	Freizeitzufriedenheit (Mittelwerte)	
Fernsehkonsument	Skala 1	Skala 2
bis 2 1/2 Stunden täglich	6.7 (27)	7.8 (17)

4.1.4 Diskussion

Diese Ergebnisse sind konsistent mit der Vermutung, daß Befragte den Kategorienvorgaben Informationen über die "reale" - oder zumindest vom Forscher vermutete/erwünschte - Verteilung entnehmen. Diese Information hilft dem Befragten zu bestimmen, was "üblich" ist, und dieser Vergleichsstandard beeinflusst sowohl den Bericht über das eigene Verhalten wie darauf bezogene Urteile.

Dementsprechend waren die Befragten eher bereit, einen hohen Fernsehkonsum zu berichten, wenn hohe Kategorienvorgaben nahelegten, daß ein solcher Konsum nicht unüblich sei, als wenn niedrigere Kategorienvorgaben nahelegten, daß "mehr als 2 1/2 Stunden" die Ausnahme seien. Darüber hinaus waren Befragte, denen die Kategorienvorgabe nahelegte, daß "andere noch viel mehr fernsehen", mit ihrer Freizeitgestaltung etwas zufriedener als Befragte, denen ein niedriger Vergleichsstandard vorgestellt wurde.

4.2 Experiment II: Einfluß kategorialer Antwortvorgaben auf nachfolgende Einstellungsmessungen

4.2.1 Spezifikation der Fragestellung

Die Überlegung, daß Kategorienvorgaben einerseits den Befragten zu einer Einschätzung dessen verhelfen, was üblich und sozial erwünscht ist, und andererseits ihr Urteil zu einer Frage Implikationen für verwandte Fragen hat, wurde in einer zweiten Untersuchung weiter verfolgt. Hierbei wurde die Kirchgangshäufigkeit der Befragten mit Kategorienvorgaben erfaßt, die entweder nahelegten, daß regelmäßiger Kirchgang üblich sei und damit die Möglichkeit boten, eigenen regelmäßigen Kirchgang zu berichten, oder nahelegten, daß Kirchgang allgemein nur sporadisch erfolge und unüblich sei. Als abhängige Variable wurden die Befragten um eine Selbsteinstufung ihrer Religiosität gebeten. Zwei Skalenversionen wurden in diesem Experiment verwendet: Version 1 war eine üblicherweise in Umfragen verwendete Kategorienabfolge, die von "mehrmals in der Woche" über "einmal in der Woche", "1-3 mal im Monat", "mehrmals im Jahr" und "seltener" bis zu "nie" reichte. Version 2 wurde zum einen umgekehrt und zum anderen extrem verkürzt dargeboten: "nie", "nur bei Hochzeiten, Todesfällen oder ähnlichen Gelegenheiten", "auch sonst 1-2 mal im Jahr" und "mehrmals im Jahr". Im Anschluß an diese Einordnungsfrage wurden die Befragten dann gebeten, ihre Religiosität auf einer 10-Punkte-Skala von 1 (nicht religiös) bis 10 (religiös) zu berichten. Skalenversion 1 legt den Befragten einerseits nahe, "daß man normalerweise in die Kirche geht" und bietet ihnen andererseits gleichzeitig die Möglichkeit, dies zu berichten; in Skalenversion 2 ist das Umgekehrte der Fall: Hier ist Kirchgang etwas Sporadisches, und die Möglichkeit auszudrücken, daß man in die Kirche geht, ist eingeschränkt.

Verschiedene Konzeptionen des Urteilsprozesses führen bei dieser experimentellen Variation zu konkurrierenden Vorhersagen. Überlegungen zum sozialen Vergleich legen nahe, daß Befragte, denen die ausführliche Skalenversion vorgelegt wird, eine höhere Kirchgangshäufigkeit des Durchschnittsbürgers erschließen als Befragte, die die verkürzte Version beantworten. Allerdings ist es unwahrscheinlich, daß diese Vergleichsinformation die berichtete Religiosität beeinflusst, da Religiosität - im Gegensatz zu Zufriedenheitsur-

teilen, wie sie in Experiment I erhoben wurden - nur in geringem Maße ein komparatives Urteil ist. Andererseits vermittelt die ausführliche Skala den Befragten, daß es üblich sei, in die Kirche zu gehen, zumindest, daß der Forscher dies annimmt. Sofern diese Information Erwägungen zur sozialen Erwünschtheit des Kirchgangs aktiviert, könnte die ausführliche Skala zum Bericht höherer Religiosität führen.

Schließlich zeigten psychologische Untersuchungen im Rahmen der Selbstwahrnehmungstheorie (z. B. BEM, 1972), daß Personen ihre Einstellungen auf der Basis ihres augenfälligen Verhaltens erschließen. Personen, die angeben, selten in die Kirche zu gehen, sollten sich daher als weniger religiös beurteilen als Personen, die angeben, häufiger in die Kirche zu gehen (vgl. SALANCIK & CONWAY, 1975). Wichtig ist in diesem Zusammenhang, daß die beiden Skalenversionen den Befragten unterschiedliche Veranlassung und Möglichkeit geben, die Häufigkeit ihres Kirchgangs zu erwägen und zu berichten. Die verkürzte Skala erlaubt Befragten, die regelmäßig die Kirche besuchen, dies nur pauschal als "mehrmals im Jahr" anzugeben. Die ausführliche Skala verlangt von diesen Befragten jedoch, die Kirchgangshäufigkeit genauer zu erinnern und bietet die Möglichkeit, differenzierte Angaben wesentlich höherer Kirchgangsfrequenzen zu geben. Dies sollte wesentliche Auswirkungen auf die nachfolgenden Urteile der Befragten über ihre eigene Religiosität haben: Eine Reihe von Untersuchungen (z. B. CARLSTON, 1980; SHERMAN et al., 1978) haben gezeigt, daß Personen, wenn sie ein Urteil abgegeben haben, bei einem nachfolgenden verwandten Urteil keine neue, umfassende Informationssuche beginnen, sondern auf das bereits getroffene Urteil zurückgreifen und die zweite Beurteilung auf der Basis der Implikationen des ersten Urteils treffen - statt auf der Basis der Information, die zum ersten Urteil führte (für eine ausführliche Darstellung siehe WYER & SRULL, 1980).

Übertragen auf das vorliegende Experiment heißt dies, daß die Befragten zur Beantwortung der Kirchgangsfrage ihr Verhalten erinnern sollten, um sich einer Kategorie zuzuordnen. Bei der nachfolgenden Beurteilung ihrer Religiosität sollten sie dann aber mit hoher Wahrscheinlichkeit keine neue Informationssuche beginnen, sondern auf den gerade abgegebenen Verhaltensbericht zurückgreifen. Dieser Verhaltensbericht wurde jedoch durch die Skalenvorgabe manipuliert: Regelmäßige Kirchgänger konnten bei der verkürzten

Skala nur angeben, "mehrmals im Jahr" in die Kirche zu gehen, während sie bei der ausführlichen Skala höhere Frequenzen berichten konnten.

Sofern regelmäßige Kirchgänger auf das erste Urteil (Verhaltensbericht) anstatt auf ihr erinnertes reales Verhalten zurückgreifen, sollten Befragte, denen die ausführliche Skala vorlag, daher höhere Religiosität berichten als Befragte, denen die verkürzte Skala vorlag. Diese Überlegung impliziert zugleich, daß Personen, die "nie" oder "selten" in die Kirche gehen, durch die Skalenunterschiede wenig beeinflusst werden sollten, da entsprechende Antwortmöglichkeiten bei beiden Skalen gegeben sind. Die Selbstwahrnehmungstheorie in Verbindung mit Ergebnissen zum verkürzten Informationsprozeß nach der Abgabe erster Urteile sagt daher Effekte der Skalenvorgabe auf nachfolgende Urteile nur für Befragte vorher, die in den Bereich "mehrmals im Jahr" fallen. Die zuvor dargestellten Überlegungen zu sozialen Vergleichsprozessen sagen hingegen Einflüsse der Skalenvorgaben auf alle Befragten vorher, was es erlaubt, diese beiden Prozesse empirisch zu unterscheiden.

4.2.2 Durchführung

Innerhalb einer telefonischen Befragung wurde 101 Personen, die angegeben hatten, einer christlichen Religionsgemeinschaft anzugehören, die Frage gestellt, "wie oft sie im allgemeinen zur Kirche gehen". Zur Beantwortung erhielten sie eine der beiden vorher beschriebenen Skalen vorgelesen. Diejenigen Befragten, die vor bzw. während des Vorlesens der Antwortkategorien eine Spontanantwort gaben, wurden von den Interviewern vermerkt und nicht in die Analyse aufgenommen. 94 Befragte verblieben in der Analyse. Als abhängige Variable wurde die Selbsteinstufung der Religiosität der Befragten mit einer 10-Punkte-Skala (1 = nicht religiös / 10 = religiös) erhoben.

4.2.3 Ergebnisse

Tabelle 7 zeigt die berichtete Kirchgangsfrequenz nach Skalenvorgabe. Wie ersichtlich, waren die Unterschiede relativ geringfügig, $\chi^2(1) = .4$, n.s.

Tab. 7: Berichtete Kirchgangsfrequenz in Abhängigkeit von der Skalenvorgabe

Skalenversion 1		Skalenversion 2		
- mehrmals in der Woche	} 48.1%	- mehrmals im Jahr	} 54.8%	
- einmal in der Woche				
- 1-3 mal im Monat				
- mehrmals im Jahr				
- seltener	} 51.9%	- nur bei Hochzeiten, Todesfällen oder ähnlichen Gelegenheiten	} 45.2%	
- nie		- auch sonst 1-2 mal im Jahr		
		- nie		
<hr/> 100% (52)		<hr/> 100% (42)		

Allerdings beeinflusste die Skalenvorgabe das nachfolgende Urteil über die eigene Religiosität, wie durch die Selbstwahrnehmungstheorie vorhergesagt (vgl. Tabelle 8).

Tab. 8: Skalenvorgabe und Selbsteinstufung der Religiosität

Kirchgangsfrequenz	Religiosität (Mittelwerte)	
	Skalenversion 1	Skalenversion 2
mehrmals im Jahr	6.5 (25)	4.8 (23)
seltener, nie	6.0 (27)	5.5 (19)
GESAMT:	6.3 (52)	5.1 (42)

Insgesamt berichteten die Befragten höhere Religiosität, wenn ihnen die ausführliche ($M = 6.3$) als wenn ihnen die verkürzte Skala ($M = 5.1$) zur Angabe der Kirchgangshäufigkeit vorlag, $F(1,91)=3.9$, $p < .06$. Wie erwartet, war dieser Effekt ausgeprägter bei Befragten, die regelmäßig in die Kirche gingen und die bei Vorlage der ausführlichen Skala die Möglichkeit hatten, hohe Kirchgangsfrequenzen zu berichten ($M = 6.5$), statt in der verkürzten Skala auf die Kategorie "mehrmals im Jahr" verwiesen zu sein ($M = 4.8$), $t(89)=2.1$, $p < .04$.

Der Einfluß der Skalenversionen auf Befragte, die selten in die Kirche gingen ($M = 6.0$ vs. 5.5) war hingegen nicht signifikant, $t(89) = .06$, n.s.

Die Ergebnisse sind somit konsistent mit der Annahme, daß die Befragten bei der Beurteilung ihrer Religiosität auf den zuvor gegebenen Verhaltensbericht zurückgriffen, dessen Inhalt durch die Art der Kategorienvorgabe manipuliert war.

4.3 Schlußbemerkung

Die beiden geschilderten Experimente verdeutlichen, daß die Breite von Kategorienvorgaben zur Erfassung quasi-objektiver Daten einerseits die Berichte der Probanden über ihr Verhalten und andererseits ihre nachfolgenden Urteile zu verwandten Fragen beeinflussen kann. Die Ergebnisse zeigen weiterhin, daß Einflüsse dieser Art im Kontext allgemeiner Theorien der sozialen Informationsverarbeitung prognostizierbar sind (zur Übersicht vgl. WYER, 1974; WYER & CARLSTON, 1979; zur Einführung: STRACK, 1983).

Diese Urteilsprozesse, die Aussagen von Befragten vorausgehen, wurden bisher von Umfrageforschern weitgehend ignoriert. Zwar verweist eine Reihe von Autoren auf den bei der Interviewsituation ablaufenden (sozialen) Interaktionsprozeß (GORDEN, 1980; CICOUREL, 1981; ESSER, 1981, 1982), doch zeichnet sich, von wenigen Ausnahmen abgesehen (etwa BISHOP et al., 1982), ein großer Teil der Auseinandersetzungen der Umfrageforschung mit Skalierungs- und Reihenfolgeeffekten durch Konzentration auf die sozialstrukturellen Bedingungen statt auf die zugrunde liegenden psychologischen Prozesse aus (SCHUMAN & PRESSER, 1981). Leider ist zu befürchten, daß die "technischen" Probleme ohne ein Verständnis der sie generierenden kognitiven Prozesse nicht gelöst werden können. Da das detaillierte Verständnis dieser Prozesse noch umfangreicher Forschung bedarf, bietet sich hier ein Forschungsgebiet an, in dem eine Zusammenarbeit von Umfrageforschern und Psychologen, die dem Paradigma der Informationsverarbeitung nahestehen, für beide Seiten fruchtbar sein könnte, da sie einerseits psychologische Forscher vor neue inhaltliche Fragen stellt und andererseits Umfrageforschern helfen könnte, der häufig verlangten "Theorie der Befragung" ein Stückchen näherzukommen. Um es mit den Worten von KALTON und SCHUMAN (1980: 38) auszudrücken: "Erst wenn wir die Faktoren, die den Befragungsprozeß bedingen, besser verstehen

und ihre Beziehungen untereinander kennen, werden wir eine Grundlage für die Konstruktion guter Fragen haben".

Der Gesamtbericht wurde von Hans-Jürgen Hippler verfaßt, der das Projekt "Befragungsexperimente" bei ZUMA betreut. An dem Teilbericht zur Fragenkonstruktion und -formulierung arbeitete Reiner Trometer (ZUMA) mit; der Untersuchungsbericht zum Einfluß kategorialer Skalenvorgaben wurde zusammen mit Norbert Schwarz (Psychologisches Institut der Universität Heidelberg) verfaßt.

Literatur

- BEM, D. J. Self-perception theory. In: L. BERKOWITZ (Eds). Advances in experimental social psychology, Vol. 6. New York: Academic Press, 1972.
- BISHOP, G. F., OLDENDICK, R. W. & TUCHFARBER, A. J. Effects of presenting one versus two sides of an issue in survey questions. Public Opinion Quarterly, 46, 1982, 69-85.
- BISHOP, G. F., OLDENDICK, R. W. & TUCHFARBER, A. J. Effects of question wording and format on political attitude consistency. Public Opinion Quarterly, 42, 1978, 81-92.
- BRADBURN, N. M. & MASON, W. N. The effect of question order on response. Journal of Marketing Research, 1, 1964, 57-61.
- CANN, A., SHERMAN, S. J. & ELKES, R. Effect of initial request size and timing of a second request on compliance: the foot in the door and the door in the face. Journal of Personality and Social Psychology, 32, 1975, 774-782.
- CANTRIL, H. (Ed.). Gauging public opinion. Princeton University Press, 1944.
- CARLSTON, D. E. The recall and use of traits and events in social inference processes. Journal of Experimental Social Psychology, 16, 1980, 303-329.
- CICOUREL, A. V. Interviewing and memory. In: C. CHERRY (Ed.). Pragmatic aspects of human communication. Dordrecht, Holland: Seidel, 1974.
- CONVERSE, P. E. Attitudes and non-attitudes: Continuation of a dialogue. In: E. R. TUFTE (Ed.). The quantitative analysis of social problems. Reading, Mass.: Addison-Wesley, 1970.
- COOK, T. D. & CAMPBELL, D. T. Quasi-experimentation. Design & analysis issues for field settings. Chicago: Rand McNally, 1979.
- DAVIS, J. A. Are surveys any good, and if so, for what? In: H. W. SINAÏKO & L. A. BROEDLING (Eds). Perspectives on attitude assessment: Surveys and their alternatives. Champaign, Ill.: Pendleton, 1976.
- DE JONG, W. An examination of self-perception mediation of the foot-in-the-door effect. Journal of Personality and Social Psychology, 37, 1979, 2221-2239.
- DIJKSTRA, W. & VAN DER ZOUWEN, J. Response behaviour in the survey-inter-

- view. London: Academic Press, 1982.
- ERBSLÖH, E., ESSER, H., RESCHKA, W. & SCHÖNE, D. Studien zum Interview. Meisenheim: Hain, 1973.
- ESSER, H. Soziale Regelmäßigkeiten des Befragtenverhaltens. Meisenheim: Hain, 1975.
- ESSER, H. Research reaction as social action and the problem of systematic measurement error. In: M. BRENNER (Ed.). Social method and social life. London: Academic Press, 1981.
- ESSER, H. Determinanten des Interviewer- und Befragtenverhaltens: Probleme der theoretischen Erklärung und empirischen Untersuchung von Interviewereffekten. In: K. U. MAYER & P. SCHMIDT (Hrsg.). Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. ZUMA-Monographien Sozialwissenschaftliche Methoden, Bd 5. Frankfurt: Campus, 1983 (im Druck).
- FERBER, R. Item nonresponse in a consumer survey. Public Opinion Quarterly, 30, 1966, 399-415.
- FRANCIS, J. D. & BUSCH, L. What we now know about "I don't know". Public Opinion Quarterly, 39, 1975, 207-218.
- FRIEDRICHS, J. Methoden empirischer Sozialforschung. Reinbek: Rowohlt, 1973.
- GORDEN, R. L. Interviewing: strategy, techniques and tactics. Homewood, Ill.: The Dorsey Press, 1980.
- HIPPLER, H.-J. & STRACK, F. Cognitive activation of values and political attitudes. Paper presented at the Fourth Annual Scientific Meeting of the International Society of Political Psychology, Mannheim, Juni 1981.
- HOLM, K. Die Befragung, Bd. 1. München: Francke, 1975.
- HYMAN, H. H. Interviewing in social research. Chicago: The University of Chicago Press, 1954.
- IRLE, M. Umfrageforschung - auch in Zukunft der "Königsweg" der empirischen Sozialforschung? Vortrag auf der gemeinsamen Tagung des Arbeitskreises Deutscher Marktforscher (ADM) und der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI), Heidelberg, Oktober 1981.
- KAHN, R. L. & CANNELL, CH. F. The dynamics of interviewing. New York: Wiley, 1957.
- KALTON, G. & SCHUMAN, H. The effect of the question on survey responses: a review. Proceedings of the Section on Survey Research Methods. Washington: American Statistical Association, 1980.
- VON KOOLWIJK, J. Fragebogenprofile. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 20, 1968, 780-791.
- VON KOOLWIJK, J. "Unangenehme Fragen". Kölner Zeitschrift für Soziologie und Sozialpsychologie, 21, 1969, 864-875.
- VON KOOLWIJK, J. & WIEKEN-MAYSER, M. (Hrsg.). Techniken der empirischen Sozialforschung, Bd. 4: Erhebungsmethoden: Die Befragung. München: Oldenbourg, 1974.
- LEVERKUS-BRÜNING, J. Die Meinungslosen. Die Bedeutung der Restkategorie in der empirischen Sozialforschung. Berlin: Duncker Humblot, 1966.
- LIPPERT, E. & WAKENHUT, R. Zur Zentralität von Einstellungen. Zeitschrift für Soziologie, Jg. 7, 1, 1978, 87-96.
- MAYNTZ, R., HOLM, K. & HÜBNER, P. Einführung in die Methoden der empiri-

- schen Soziologie. Opladen: Westdeutscher Verlag, 1971.
- NOELLE, E. Umfragen in der Massengesellschaft. Hamburg: Rowohlt, 1963.
- NOELLE-NEUMANN, E. Wanted: Rules for wording structured questionnaires. Public Opinion Quarterly, 34, 1970, 191-201.
- RUGG, D. Experiments in wording questions: II. Public Opinion Quarterly, 5, 1941, 91-92.
- SALANCIK, J. R. & CONWAY, M. Attitude inferences from salient and relevant cognitive content about behavior. Journal of Personality and Social Psychology, 32, 1975, 829-842.
- SCHEUCH, E. K. Das Interview in der Sozialforschung. In: R. KÖNIG (Hrsg.). Handbuch der empirischen Sozialforschung. Stuttgart: Enke, 1973.
- SCHUMAN, H. & PRESSER, S. Question wording as an independent variable in survey analysis. Sociological Methods and Research, 6, 1977, 151-170.
- SCHUMAN, H. & PRESSER, S. The assessment of "no opinion" in attitude surveys. In: K. F. SCHUESLER (Ed.). Sociological Methodology. San Francisco: Jossey-Bass, 1979.
- SCHUMAN, H. & PRESSER, S. Question answers in attitude surveys. Experiments on question form, wording, and context. New York: Academic Press, 1981.
- SCHWARZ, N. & STRACK, F. Kognitive und affektive Prozesse bei der Beurteilung des subjektiven Wohlbefindens. Projektantrag. Heidelberg/Mannheim, 1982.
- SCHWARZ, N., WYER, R. S. & KRUGLANSKI, A. W. Effekte von Rangordnungsaufgaben auf nachfolgende Bewertungen. Vortrag bei der 24. Tagung experimentell arbeitender Psychologen, Trier, April 1982, auch MAFFIS-Bericht Nr. 1, 1982.
- SHERMAN, S. J., AHLM, K., BERMAN, L. & LYNN, S. Contrast effects and their relationship to subsequent behavior. Journal of Experimental Social Psychology, 14, 1978, 430-450.
- SNYDER, M. & CUNNINGHAM, M. R. To comply or not comply: Testing the self-perception explanation of the foot-in-the-door phenomenon. Journal of Personality and Social Psychology, 31, 1975, 64-67.
- STRACK, F. Soziale Informationsverarbeitung. In: D. FREY & S. GREIF (Hrsg.). Sozialpsychologie - Ein Handbuch in Schlüsselbegriffen. München: Urban & Schwarzenberg, 1983 (im Druck).
- SUDMAN, S. & BRADBURN, N. M. Asking questions. San Francisco: Jossey-Bass, 1982.
- TVERSKY, A. & KAHNEMAN, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 1973, 207-232.
- UPSHAW, H. S. The personal reference scale. In: L. BERKOWITZ (Ed.). Advances in experimental social psychology, Vol. 4. New York: Academic Press, 1969.
- WYER, R. S. Cognitive organization and change: An information-processing approach. Potomac, Maryland: Erlbaum, 1974.
- WYER, R. S. & CARLSTON, D. E. Social cognition, inference, and attribution. Hillsdale, N.J.: Erlbaum, 1979.
- WYER, R. S. & SRULL, T. K. The processing of social information: A conceptual integration. In: R. HASTIE et al. (Eds.). Person memory. Hillsdale, N.J.: Erlbaum, 1980.